Semiparametrically efficient estimation of Euclidean parameters under equality constraints

Klaassen, C.A.J.; Susyanto, N.

DOI
10.1016/j.jspi.2018.12.005
Publication date
2019
Document Version
Final published version
Published in
Journal of Statistical Planning and Inference

Journal of Statistical Planning and Inference 201 (2019) 120–132

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Semiparametrically efficient estimation of Euclidean parameters under equality constraints

Chris A.J. Klaassen [a], Nanang Susyanto [b,*]

[a] *Korteweg–de Vries Institute for Mathematics, University of Amsterdam, The Netherlands*
[b] *Department of Mathematics, Universitas Gadjah Mada, Yogyakarta, Indonesia*

ARTICLE INFO

Article history:
Received 20 March 2018
Received in revised form 2 November 2018
Accepted 16 December 2018
Available online 26 December 2018

MSC:
62F30
62F12
62F10

Keywords:
Semiparametric estimation
Semiparametric submodels
Efficient estimator
Equality restrictions
Functional equality

ABSTRACT

Assume a (semi)parametrically efficient estimator is given of the Euclidean parameter in a (semi)parametric model. A submodel is obtained by constraining this model in that a continuously differentiable function of the Euclidean parameter vanishes. We present an explicit method to construct (semi)parametrically efficient estimators of the Euclidean parameter in such equality constrained submodels and prove their efficiency. Our construction is based solely on the original efficient estimator and the constraining function.

1. Introduction

There is an extensive literature about *estimation under constraints*, mainly of finite-dimensional parameters. Density estimation under shape restrictions, like monotonicity and convexity, may be viewed as estimation under constraints of infinite-dimensional parameters. See e.g. Groeneboom et al. (2001) and Horowitz and Lee (2017).

We do not consider nonparametric estimation and restrict attention to the estimation of Euclidean parameters within parametric and semiparametric models. Let us first consider parametric models with their natural parameter space and assume that this natural parameter is constrained to a subset of the parameter space and that this subset has a nonempty interior. Typically such subsets are defined by inequalities for the parameter. So-called restricted maximum likelihood estimators have been studied for many such models, and admissible and minimax estimators have been constructed as well, also in the presence of nuisance parameters. An excellent overview of the huge literature on this topic is given in the monograph by Van Eeden (2006) with 27 pages of references. A more concise review is presented by Marchand and Strawderman (2004).

When we study these estimation problems from the point of view of asymptotic statistics, Local Asymptotic Normality plays an important role. Local Asymptotic Normality at a particular parameter value implies that the experiment of estimating the parameter in the neighborhood of the particular value when properly rescaled, as sample size tends to infinity, converges to the experiment of estimating the location parameter of a (multivariate) normal distribution based on one

* Corresponding author.
E-mail addresses: c.a.j.klaassen@uva.nl (C.A.J. Klaassen), nanang_susyanto@ugm.ac.id (N. Susyanto).

https://doi.org/10.1016/j.jspi.2018.12.005
0378-3758/

observation, a so-called Gaussian shift experiment. A generalization of this insight stems from Lucien Le Cam; cf. Le Cam (1972). See also Chapter 10 of Cam (1986) and Chapter 9 of Van der Vaart (1998). Consider regular parametric models, as defined e.g. in Section 2.1 of Bickel et al. (1993). The regularity implies that the parameter space is open and that Local Asymptotic Normality holds at every parameter value. Consequently, asymptotically efficient estimators of the parameter exist; see e.g. Section 2.5 of Bickel et al. (1993). As asymptotically efficient estimators are consistent and hence take their values in any neighborhood however small of the true parameter value with probability tending to 1 as sample size increases, any estimator that is efficient within the original model, is efficient in the constrained model as well, at least if the parameter space of the constrained model is again open. In case the restricted parameter set is closed, one has to work with tangent cones in order to study efficient estimation for parameter values at the boundary; see Chapters 2 and 5 of Van der Vaart (1998), and Andrews (1999).

Next, consider semiparametric estimation with restrictions or constraints on the Euclidean parameter that result in parameter spaces with nonempty interior. Here too, for parameter values in the interior typically Local Asymptotic Normality holds, which means that the standard asymptotic efficiency theory holds, i.e., the sequence of experiments indexed by the sample size behaves asymptotically like the experiment of one observation from a normal distribution for which the mean vector has to be estimated and for which there are no restrictions on this mean vector. Again, for parameter values at the boundary of the restricted subset, there are restrictions on the mean vector in the limit experiment, which typically amount to the mean vector being restricted to a cone.

However, if the restrictions on the parameter space of a parametric or semiparametric model lead to a parameter subset with empty interior, asymptotic estimation theory becomes interesting again. If the restrictions lead to a subset that can be reparametrized in such a way that we are dealing with a regular parametric problem, then Susyanto and Klaassen (2017) can be applied. However, there are constraints that result in parameter subsets for which such a reparametrization is impossible. For example, consider bivariate normal distributions for which the mean vectors are known to be on the unit circle, or any other closed curve without self-intersections. As such a closed curve cannot be reparametrized via an open interval, and hence the constrained model cannot be viewed as a regular parametric model, Susyanto and Klaassen (2017) cannot be applied and a different approach has to be developed. This is done in the present paper, which considers parameter subsets that are zero sets of continuous (constraining) functions and consequently closed. On the other hand, stereographic projection of the unit circle from its north pole shows that the unit circle with one point deleted can be parametrized via an open set. However, the unit circle with one point deleted is not closed. This means that the present paper and Susyanto and Klaassen (2017) are complementary in some sense; although many models can be treated by the methods of both of them, there are models that can be treated by one but not the other.

In the present paper, the asymptotic lower bound for the constrained model is obtained by projection of the efficient influence function within the unconstrained model, and an efficient estimator for the constrained model is constructed by correcting the original, efficient estimator for the unconstrained model with an expression involving the derivative of the constraining function at the value of the original estimator. In contrast, Susyanto and Klaassen (2017) obtain the asymptotic lower bound for the constrained model via projection of efficient score functions and they start from a $\sqrt{n}$-consistent estimator of the underlying parameter in the constrained model, correct it with an expression involving the given efficient estimator for the unconstrained model, and apply the reparametrization function to it in order to obtain their efficient estimator for the constrained model. In the present paper, everything will be done directly to the original parameter subject to equality constraints without reparametrizing it. Despite the fundamental differences between the present paper and Susyanto and Klaassen (2017), their methods yield the same estimator in case the constraining function is linear.

The outline of the paper is as follows. The type of models we consider and the necessary notation will be introduced in Section 2. There we discuss some more of the literature on constrained models, especially on constrained *nonparametric* models. In Section 3, we will present a lower bound to the efficient information bound for estimating the parameter of interest within the constrained model. This lower bound will be formulated in terms of the efficient information bound of the original, unconstrained model and the Jacobian of the constraining function. An explicit estimator that is efficient within the constrained model, will be given in Section 4. It attains the lower bound from Section 3, which shows that both this information bound and the estimator are efficient within the constrained model. Examples are discussed in Section 5. Our conclusions are presented in Section 6.

2. Model

With Θ an open subset of $\mathbb{R}^k$, let

$$\mathcal{P} = \left\{ P_{\theta,G} \ : \ \theta \in \Theta, \ G \in \mathcal{G} \right\} \tag{2.1}$$

be a collection of distributions on some measurable space $(\mathcal{X}, \mathcal{A})$ indexed by θ and G. Here, θ is the parameter of interest and G is the nuisance parameter. Typically, the nuisance parameter space $\mathcal{G}$ is a subset of a Banach or Hilbert space and $\mathcal{P}$ represents a semiparametric model then. If $\mathcal{G}$ is finite-dimensional, we are dealing with a parametric model. Let the random variables $X_1, \ldots, X_n$ be i.i.d. copies of X taking values in $\mathcal{X}$ under distribution $P_{\theta,G} \in \mathcal{P}$. The model $\mathcal{P}$ is understood to be regular, i.e., for every $G \in \mathcal{G}$ the parametric submodel $\left\{ P_{\theta,G} \ : \ \theta \in \Theta \right\}$ is regular.

We assume an estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ of the parameter of interest θ is given that is asymptotically linear in the efficient influence function $\tilde{\ell}(\cdot; \theta, G, \mathcal{P})$, which means that

$$\sqrt{n}\left(\hat{\theta}_n - \theta - \frac{1}{n}\sum_{i=1}^{n} \tilde{\ell}(X_i; \theta, G, \mathcal{P})\right) \rightarrow_{P_{\theta,G}} 0, \quad \theta \in \Theta, \quad G \in \mathcal{G}, \tag{2.2}$$

holds. Such an estimator is called asymptotically efficient in estimating θ within $\mathcal{P}$ because it attains the asymptotic information bound

$$I^{-1}(\theta, G, \mathcal{P}) = \int_{\mathcal{X}} \tilde{\ell}(x; \theta, G, \mathcal{P})\tilde{\ell}^T(x; \theta, G, \mathcal{P})dP_{\theta,G}(x), \tag{2.3}$$

which corresponds to the efficient information matrix $I(\theta, G, \mathcal{P})$.

Quite frequently the elements of the parameter of interest $\theta = (\theta_1, \ldots, \theta_k)$ are not mathematically independent but satisfy d functional relationships $S_i(\theta) = 0,\ i = 1, \ldots, d$, with $d < k$. Formally, this can be described as

$$S(\theta) = 0, \quad \theta \in \Theta, \tag{2.4}$$

where S is a function from $\mathbb{R}^k$ to $\mathbb{R}^d$. We will assume that the $d \times k$ Jacobian matrix $\dot{S}(\cdot)$ exists, is continuous in θ on Θ, and has full rank d. Thus, we have constrained the semiparametric model $\mathcal{P}$ to a semiparametric submodel of it, namely

$$\mathcal{Q} = \left\{P_{\theta,G} \,:\, S(\theta) = 0,\ \theta \in \Theta,\ G \in \mathcal{G}\right\}. \tag{2.5}$$

Given the constraint $S(\theta) = 0$, we will adapt the semiparametrically efficient estimator $\hat{\theta}_n$ of θ within $\mathcal{P}$ in such a way that the adapted estimator is semiparametrically efficient within the constrained model $\mathcal{Q}$. As one has more information ($S(\theta) = 0$) available about θ in the constrained model $\mathcal{Q}$ than in the unconstrained model $\mathcal{P}$, the adapted estimator has to have at least as small asymptotic variance as the original estimator $\hat{\theta}_n$ and to be at least as close to the true value stochastically.

Efficient estimation of Euclidean parameters under equality constraints for *nonparametric* models has been studied in Levit (1976), Koshevnik and Levit (1977) and Haberman (1984), in Example 1.3.6, 3.2.3, and 3.3.3 of Bickel et al. (1993), in Müller and Wefelmeyer (2002), and in Broniatowski and Keziou (2012). In Bickel et al. (1993) nonparametric models under equality constraints are called constraint defined models. Let the semiparametric model $\mathcal{P}$ be embedded into a nonparametric model $\tilde{\mathcal{P}}$ and let the map $\nu : \tilde{\mathcal{P}} \rightarrow \mathbb{R}^k$ be such that $\nu(P_{\theta,G}) = \theta$ holds for all $P_{\theta,G} \in \mathcal{P}$. In view of $\mathcal{P} \subset \tilde{\mathcal{P}}$ estimation of $\nu(P)$ within $\mathcal{P}$ is easier than within $\tilde{\mathcal{P}}$. This relation between these models also holds under the equality constraint $S(\nu(P)) = 0$. Consequently, the results for nonparametric models under constraints are not directly applicable to our semiparametric situation.

For the constrained parametric estimation problem so-called restricted maximum likelihood estimators have been studied. In Aitchison and Silvey (1958) this has been done with an iterative computation method using Lagrange multipliers. An alternative iterative construction has been proposed in Jamshidian (2004), which also presents a long list of examples of constrained parametric estimation problems. To prove efficiency of these restricted maximum likelihood estimators additional regularity conditions are needed. Our method does not need these additional conditions, provided an efficient estimator for the original unconstrained parametric model is given. Finite sample Cramér–Rao bounds for the constrained parametric case have been derived in e.g. Gorman and Hero (1990), Marzetta (1993), and Stoica and Ng (1998).

To the best of our knowledge the semiparametric version of the topic of the present paper has not been studied in literature yet, apart from Susyanto and Klaassen (2017). The differences between this paper and the present one are explained in Section 1.

If $\mathcal{Q}$ can be reparametrized as

$$\mathcal{Q} = \left\{P_{f(\nu),G} \,:\, \nu \in N,\ G \in \mathcal{G}\right\}, \tag{2.6}$$

where N is open and $f : N \rightarrow \Theta$ is injective and continuously differentiable with full rank Jacobian, then ν can be estimated semiparametrically efficiently as in Susyanto and Klaassen (2017) and, as noted there, θ can be estimated efficiently as well by applying $f(\cdot)$ to the efficient estimator of ν. However, it may be hard or even impossible to find such a reparametrization. As mentioned already in Section 1, a simple, formal example of this impossibility is estimation of the mean vector of a bivariate normal distribution where it is known that this mean vector lies on the unit circle; see Example 5.1. The unit circle cannot be parametrized as in (2.6) with N open and $f(\cdot)$ continuous and injective. Indeed, assume $\nu_n \in N$ converge to a point at the boundary of N. Then $f(\nu_n)$ converge to a point on the unit circle $f(\nu_0)$, say, with $\nu_0 \in N$. But by the continuity of $f(\cdot)$ this implies that there exist a point in N close to the boundary of N and another point in N close to ν_0 that are mapped on the same point of the circle by $f(\cdot)$, which contradicts its injectivity. On the other hand there are submodels $\mathcal{Q}$ of the type (2.6) that cannot be viewed as a submodel of the type (2.5), as mentioned already in Section 1. Again consider estimation of the mean vector of a bivariate normal distribution where it is known now that this mean vector lies on the unit circle with one point removed. This unit circle with one point removed can be parametrized as in (2.6) with $f(\cdot)$ continuous and N open, but it cannot be described via (2.4), since the preimage of the closed set $\{0\}$ under a continuous function $S(\cdot)$ has to be closed and the unit circle with one point removed is not.

3. Efficient influence functions and projection

In the situation of Section 1 we denote the so-called *efficient score function* for θ by

$$\ell^*(\cdot;\theta,G,\mathcal{P})=I(\theta,G,\mathcal{P})\tilde{\ell}(\cdot;\theta,G,\mathcal{P}). \tag{3.1}$$

We will restrict attention to regular semiparametric models for which at every $P_0=P_{\theta_0,G_0}\in\mathcal{P}$ the parameter θ is pathwise differentiable, the tangent space $\dot{\mathcal{P}}$ is the sum of the tangent space $\dot{\mathcal{P}}_1$ for θ and the tangent space $\dot{\mathcal{P}}_2$ for G, and the efficient score function $\ell^*(\cdot;\theta,G,\mathcal{P})$ for θ is the projection of the (ordinary) score function $\dot{\ell}(\cdot;\theta,G,\mathcal{P})$ for θ on the orthocomplement of $\dot{\mathcal{P}}_2$ within $\dot{\mathcal{P}}$ in the sense of componentwise projection within $L_2^0(P_0)=\{f\in L_2(P_0):E_{P_0}f(X)=0\}$; for details see Chapter 3 of Bickel et al. (1993) or Chapter 25 of Van der Vaart (1998).

By Proposition 3.3.1 of Bickel et al. (1993) the efficient influence function $\tilde{\ell}(\cdot;\theta,G,\mathcal{Q})$ for θ within the submodel $\mathcal{Q}$ can be obtained by projecting the efficient influence function $\tilde{\ell}(\cdot;\theta,G,\mathcal{P})$ for θ within $\mathcal{P}$ onto the tangent space $\dot{\mathcal{Q}}$ of $\mathcal{Q}$ or onto an appropriate subspace of this tangent space.

Let $\{\theta_\eta\ :\ \theta_\eta\in\mathbb{R}^k,\ \eta\in\mathbb{R},\ |\eta|<\epsilon\}$ for sufficiently small $\epsilon>0$ be a path through θ_0 in $\mathbb{R}^k$ in the direction $r\in\mathbb{R}^k$, which means that $|\theta_\eta-\theta_0-\eta r|=o(|\eta|)$. If this path satisfies $S(\theta_\eta)=0,\ |\eta|<\epsilon$, then the differentiability of $S(\cdot)$ at θ_0 implies $|S(\theta_\eta)-S(\theta_0)-\eta\dot{S}(\theta_0)r|=o(|\eta|)$, meaning $|\eta\dot{S}(\theta_0)r|=o(|\eta|)$, and hence $\dot{S}(\theta_0)r=0$. In other words, such a path within the parameter set $\{\theta\ :\ S(\theta)=0,\ \theta\in\mathbb{R}^k\}$, has a direction r at θ_0 that belongs to the orthocomplement of the d-dimensional linear space within $\mathbb{R}^k$ spanned by the d row vectors of the Jacobian matrix $\dot{S}(\theta_0)$. In fact, to each element of this orthocomplement $[\dot{S}(\theta_0)]^\perp$ corresponds such a path, as is proved in detail in Appendix with the help of the implicit function theorem.

With $P_0\in\mathcal{Q}$ let L be a $k\times(k-d)$-matrix, whose columns span this $(k-d)$-dimensional orthocomplement. Since $\mathcal{P}$ is a regular semiparametric model, the parametric submodel $\mathcal{P}_1=\{P_{\theta,G_0}\ :\ \theta\in\Theta\}$ is regular. With $s(\theta)$ denoting the square root of the density of P_{θ,G_0} with respect to an appropriate dominating measure μ, this regularity implies

$$\|s(\theta_\eta)-s(\theta_0)-\frac{1}{2}s(\theta_0)(\theta_\eta-\theta_0)^T\dot{\ell}(\theta_0)\|_\mu=o(|\theta_\eta-\theta_0|),\quad \theta_\eta\to\theta_0, \tag{3.2}$$

where $\|\cdot\|_\mu$ is the norm of $L_2(\mu)$ and $\dot{\ell}(\theta_0)=\dot{\ell}(\cdot;\theta_0,G_0,\mathcal{P})$ is the score function for θ at θ_0; cf. Definition 2.1.1 and formula (2.1.4) of Bickel et al. (1993). For a path $\{\theta_\eta\ :\ \theta_\eta\in\mathbb{R}^k,\ \eta\in\mathbb{R},\ |\eta|<\epsilon\}$ with direction r at θ_0 as above, this implies

$$\|s(\theta_\eta)-s(\theta_0)-\frac{1}{2}\eta s(\theta_0)r^T\dot{\ell}(\theta_0)\|_\mu=o(|\eta|),\quad \eta\to 0. \tag{3.3}$$

Consequently, we are dealing here with a 1-dimensional regular parametric model with score function $r^T\dot{\ell}(\theta_0)$ for η at $\eta=0$. It follows that the closed linear span $\left[L^T\dot{\ell}(\theta_0)\right]$ of all such score functions $r^T\dot{\ell}(\theta_0)$ is the tangent space $\dot{\mathcal{Q}}_1$ of $\mathcal{Q}_1=\{P_{\theta,G_0}\ :\ S(\theta)=0,\ \theta\in\Theta\}$ at P_0. This implies that the tangent space $\dot{\mathcal{Q}}$ of $\mathcal{Q}$ at P_0 contains both $\left[L^T\dot{\ell}(\theta_0)\right]$ and $\dot{\mathcal{P}}_2$. Writing $\ell^*(\theta_0)$ for $\ell^*(\cdot;\theta_0,G_0,\mathcal{P})$ we have for every tangent $t\in\dot{\mathcal{P}}_2$

$$r^T\dot{\ell}(\theta_0)+t=r^T\ell^*(\theta_0)+t+r^T\left(\dot{\ell}(\theta_0)-\ell^*(\theta_0)\right). \tag{3.4}$$

Since $\ell^*(\theta_0)$ is the componentwise projection of $\dot{\ell}(\theta_0)$ on the orthocomplement of $\dot{\mathcal{P}}_2$, each component of $\dot{\ell}(\theta_0)-\ell^*(\theta_0)$ belongs to $\dot{\mathcal{P}}_2$ and we obtain from (3.4)

$$\dot{\mathcal{Q}}\supset\left[L^T\dot{\ell}(\theta_0)\right]+\dot{\mathcal{P}}_2=\left[L^T\ell^*(\theta_0)\right]+\dot{\mathcal{P}}_2\supset\left[L^T\ell^*(\theta_0)\right]. \tag{3.5}$$

Taking $\theta=\theta_0$ in formula (3.1) and suppressing θ_0 and $\mathcal{P}$ from the notation, we rewrite (3.5) as

$$\dot{\mathcal{Q}}\supset\left[L^T\dot{\ell}\right]+\dot{\mathcal{P}}_2=\left[L^T\ell^*\right]+\dot{\mathcal{P}}_2\supset\left[L^T\ell^*\right]=\left[L^TI\tilde{\ell}\right]. \tag{3.6}$$

We shall denote the componentwise inner product within $L_2^0(P_0)$ by $\langle\cdot,\cdot\rangle_0$ and the projection within $L_2^0(P_0)$ of the efficient influence function $\tilde{\ell}$ into $\left[L^TI\tilde{\ell}\right]$ by

$$\Pi_0\left(\tilde{\ell}\mid\left[L^TI\tilde{\ell}\right]\right)=AL^TI\tilde{\ell}, \tag{3.7}$$

where A is a $k\times(k-d)$-matrix. Since $\tilde{\ell}-\Pi_0\left(\tilde{\ell}\mid\left[L^TI\tilde{\ell}\right]\right)$ has to be orthogonal to $\left[L^TI\tilde{\ell}\right]$, i.e., since

$$\left\langle\tilde{\ell}-AL^TI\tilde{\ell}\,,\,\tilde{\ell}^TIL\right\rangle_0=I^{-1}IL-AL^TII^{-1}IL=0 \tag{3.8}$$

holds, we have

$$\Pi_0\left(\tilde{\ell}\mid\left[L^TI\tilde{\ell}\right]\right)=L\left(L^TIL\right)^{-1}L^TI\tilde{\ell}. \tag{3.9}$$

In order to write this projection in terms of $\dot{S}=\dot{S}(\theta_0)$ we note that according to the Appendix of Susyanto and Klaassen (2017) $L(L^TIL)^{-1}L^TI+I^{-1}\dot{S}^T(\dot{S}I^{-1}\dot{S}^T)^{-1}\dot{S}$ is the identity map, which implies

$$\Pi_0\left(\tilde{\ell}\mid\left[L^TI\tilde{\ell}\right]\right)=\tilde{\ell}-I^{-1}\dot{S}^T(\dot{S}I^{-1}\dot{S}^T)^{-1}\dot{S}\tilde{\ell}. \tag{3.10}$$

By Theorem 3.3.2.A of Bickel et al. (1993) and formula (3.3.27) in particular, this implies that the limit distribution under P_0 of any properly normalized regular estimator of θ within the submodel $\mathcal{Q}$ is the convolution of a normal distribution with mean 0 and covariance matrix

$$L\left(L^T I L\right)^{-1} L^T = I^{-1} - I^{-1}\dot{S}^T(\dot{S}I^{-1}\dot{S}^T)^{-1}\dot{S}I^{-1} \tag{3.11}$$

and some other distribution. In the next section we shall construct an estimator $\tilde{\theta}$ of θ within $\mathcal{Q}$ that is asymptotically linear in the influence function from (3.10). Consequently, it is asymptotically normal with minimal covariance matrix, i.e.,

$$\sqrt{n}\left(\tilde{\theta} - \theta_0\right) \to_{P_0} \mathcal{N}\left(0, I^{-1} - I^{-1}\dot{S}^T(\dot{S}I^{-1}\dot{S}^T)^{-1}\dot{S}I^{-1}\right) \tag{3.12}$$

holds.

4. Efficient estimator under equality constraints

Note that $S(\theta) = S(\theta) - S(\theta_0) = \dot{S}(\theta_0)(\theta - \theta_0) + o(|\theta - \theta_0|)$ holds for θ_0 with $S(\theta_0) = 0$. Since an efficient estimator $\hat{\theta}_n$ within $\mathcal{P}$ is asymptotically linear in the efficient influence function $\tilde{\ell}(\cdot;\theta, G, \mathcal{P})$, this implies that $S(\hat{\theta}_n)$ is asymptotically linear in the influence function $\dot{S}(\theta_0)\tilde{\ell}(\cdot;\theta_0, G_0, \mathcal{P})$ under θ_0. In order to construct an efficient estimator of θ within $\mathcal{Q}$ we will use this asymptotic linearity.

Our main result reads as follows.

Theorem 4.1. *Consider the regular semiparametric model $\mathcal{P}$ and its submodel $\mathcal{Q}$ given by (2.1) and (2.5), respectively. Assume that $S : \mathbb{R}^k \to \mathbb{R}^d$, $d < k$, is continuously differentiable with Jacobian matrix $\dot{S}(\cdot)$ of full rank d, and that the tangent spaces satisfy the conditions mentioned in the first paragraph of Section 3. Let $X_1, \dots, X_n$ be i.i.d. with distribution $P_{\theta,G} \in \mathcal{P}$ and suppose that $\hat{\theta}_n$ is an efficient estimator of the parameter of interest θ within $\mathcal{P}$ based on $X_1, \dots, X_n$ with efficient influence function $\tilde{\ell}(\cdot;\theta, G, \mathcal{P})$ and that $\hat{I}_n$ is a consistent estimator of $I(\theta, G, \mathcal{P})$ from (2.3). Write*

$$\theta_n^* = \hat{\theta}_n - \hat{I}_n^{-1}\dot{S}^T(\hat{\theta}_n)\left(\dot{S}(\hat{\theta}_n)\hat{I}_n^{-1}\dot{S}^T(\hat{\theta}_n)\right)^{-1} S(\hat{\theta}_n) \tag{4.1}$$

and define

$$\tilde{\theta}_n = \mathop{\arg\min}_{\zeta,\, S(\zeta)=0} \|\zeta - \theta_n^*\| \tag{4.2}$$

with $\|\cdot\|$ the Euclidean norm or a topologically equivalent norm. Then $\tilde{\theta}_n$ efficiently estimates θ within the submodel $\mathcal{Q}$ with efficient influence function

$$\begin{aligned}\tilde{\ell}(\cdot;\theta, G, \mathcal{Q}) = {} & \tilde{\ell}(\cdot;\theta, G, \mathcal{P}) \\ & - I^{-1}(\theta, G, \mathcal{P})\dot{S}^T(\theta)\left(\dot{S}(\theta)I^{-1}(\theta, G, \mathcal{P})\dot{S}^T(\theta)\right)^{-1}\dot{S}(\theta)\tilde{\ell}(\cdot;\theta, G, \mathcal{P})\end{aligned} \tag{4.3}$$

and hence it satisfies (3.12). Furthermore,

$$\sqrt{n}(\tilde{\theta}_n - \theta_n^*) \to_{P_{\theta,G}} 0 \tag{4.4}$$

holds with $P_{\theta,G} \in \mathcal{Q}$.

Proof. In view of the convolution result proved in Section 3 (cf. (3.11)) it suffices to show that $\tilde{\theta}_n$ is asymptotically linear in the influence function from (4.3), since this yields both sharpness of the convolution bound and efficiency of the estimator. Fix θ_0 with $S(\theta_0) = 0$ and $P_0 = P_{\theta_0,G_0} \in \mathcal{Q}$, and write

$$\begin{aligned}&\sqrt{n}\left(\theta_n^* - \theta_0 - \frac{1}{n}\sum_{i=1}^n \tilde{\ell}(X_i;\theta_0, G_0, \mathcal{Q})\right) \\ &\quad = \sqrt{n}\left(\hat{\theta}_n - \theta_0 - \frac{1}{n}\sum_{i=1}^n \tilde{\ell}(X_i;\theta_0, G_0, \mathcal{P})\right) \\ &\quad\quad - \hat{I}_n^{-1}\dot{S}^T(\hat{\theta}_n)\left(\dot{S}(\hat{\theta}_n)\hat{I}_n^{-1}\dot{S}^T(\hat{\theta}_n)\right)^{-1} \\ &\quad\quad\quad \times\sqrt{n}\left(S(\hat{\theta}_n) - \frac{1}{n}\sum_{i=1}^n \dot{S}(\theta_0)\tilde{\ell}(X_i;\theta_0, G_0, \mathcal{P})\right) \\ &\quad\quad - \left(\hat{I}_n^{-1}\dot{S}^T(\hat{\theta}_n)\left(\dot{S}(\hat{\theta}_n)\hat{I}_n^{-1}\dot{S}^T(\hat{\theta}_n)\right)^{-1}\right. \\ &\quad\quad\quad \left. - I^{-1}(\theta_0, G_0, \mathcal{P})\dot{S}^T(\theta_0)\left(\dot{S}(\theta_0)I^{-1}(\theta_0, G_0, \mathcal{P})\dot{S}^T(\theta_0)\right)^{-1}\right)\dot{S}(\theta_0)\end{aligned} \tag{4.5}$$

$$\times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\ell}(X_i; \theta_0, G_0, \mathcal{P})$$
$$= R_{1,n} - R_{2,n} - R_{3,n}.$$

The asymptotic linearity of $\hat{\theta}_n$ from (2.2) implies that $R_{1,n}$ converges to 0 in probability under P_0. By the central limit theorem the second factor of $R_{3,n}$ is asymptotically normal with mean 0 and covariance matrix $I^{-1}(\theta_0, G_0, \mathcal{P})$ from (2.3). Since $\dot{S}(\cdot)$ is continuous and $\hat{I}_n$ and $\hat{\theta}_n$ are consistent in estimating $I(\theta_0, G_0, \mathcal{P})$ and θ_0, respectively, this implies that $R_{3,n}$ converges to 0 in probability under P_0 as well. We also conclude that the first factor of $R_{2,n}$ is bounded in probability. Together with the asymptotic linearity of $S(\hat{\theta}_n)$, as noted at the start of this section, this yields the convergence of $R_{2,n}$ to 0 in probability under P_0.

It remains to be shown that (4.4) holds. In view of $S(\theta_0) = 0$ and Appendix we may parametrize a part of the zero set of $S(\cdot)$ near θ_0 by

$$\mathcal{S}_0 = \{\theta \mid \theta = \theta_0 + L\eta + r(\eta),\ \eta \in H\}, \tag{4.6}$$

where the $k - d$ columns of the matrix L span the orthocomplement of $[\dot{S}(\theta_0)]$, $r(\eta) = o(\|\eta\|)$ holds as $\|\eta\|$ tends to 0, and H is an appropriate neighborhood of 0 within $\mathbb{R}^{k-d}$. The expectation of an efficient influence function vanishes, in particular, $E_{P_0}(\tilde{\ell}(X; \theta_0, G_0, \mathcal{Q})) = 0$. Consequently, the central limit theorem implies that $n^{-1} \sum_{i=1}^{n} \tilde{\ell}(X_i; \theta_0, G_0, \mathcal{Q})$ is of the order $O_p(1/\sqrt{n})$ under P_0. Furthermore, it takes its values in $[L]$ in view of (3.9). Together with (4.6) this shows that there exists a random k-vector $\tilde{R}_n = o_p(1/\sqrt{n})$ such that

$$\theta_0 + \frac{1}{n} \sum_{i=1}^{n} \tilde{\ell}(X_i; \theta_0, G_0, \mathcal{Q}) + \tilde{R}_n \in \mathcal{S}_0 \tag{4.7}$$

holds with probability tending to 1. Because of the definition of $\tilde{\theta}_n$, the triangle inequality, and the asymptotic linearity of θ_n^* in the efficient influence function as proved above (cf. (4.5)), this yields

$$\begin{aligned} &\|\tilde{\theta}_n - \theta_n^*\| \\ &\quad \le \|\theta_0 + \frac{1}{n} \sum_{i=1}^{n} \tilde{\ell}(X_i; \theta_0, G_0, \mathcal{Q}) + \tilde{R}_n - \theta_n^*\| \\ &\quad \le \|\theta_0 + \frac{1}{n} \sum_{i=1}^{n} \tilde{\ell}(X_i; \theta_0, G_0, \mathcal{Q}) - \theta_n^*\| + \|\tilde{R}_n\| \\ &\quad = o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{4.8}$$

which proves (4.4).

Remark 4.1. Theorem 4.1 states properties of the behavior of $\tilde{\theta}_n$ under $P_{\theta,G} \in \mathcal{Q}$, its proof shows that for these properties to hold only properties of the behavior of $\hat{\theta}_n$ and $\hat{I}_n$ under $P_{\theta,G} \in \mathcal{Q}$ are needed. In particular, this means that $\hat{I}_n$ has to converge in probability to $I(\theta, G, \mathcal{P})$ under $P_{\theta,G}$ only for θ with $S(\theta) = 0$. A similar remark holds for Theorem 4.1 of Susyanto and Klaassen (2017). In fact this has been used in the Gaussian copula Example 5.3 of ibid. where $\hat{I}_n$ estimates $I(f(\rho), G, \mathcal{P})$.

Remark 4.2. Consistent estimators $\hat{I}_n$ of $I(\theta, G, \mathcal{P})$ may be constructed from $\hat{\theta}_n$ as in Section 4 of Susyanto and Klaassen (2017). In regular parametric cases the Fisher information $I(\theta) = I(\theta, G, \mathcal{P})$ depends on θ only and is continuous in it. Consequently, $\hat{I}_n = I(\hat{\theta}_n)$ is consistent in estimating $I(\theta)$ then.

Remark 4.3. According to Theorem 4.1 the estimators θ_n^* and $\tilde{\theta}_n$ have the same asymptotic performance to first order. However, only $\tilde{\theta}_n$ is guaranteed to be efficient within $\mathcal{Q}$, as θ_n^* need not be a zero of $S(\cdot)$. In order to compute $\tilde{\theta}_n$ to the desired order of precision one typically needs an iterative numerical procedure, like Newton–Raphson.

Remark 4.4. Parametrize the linear case by $S(\theta) = R^T(\theta - \alpha)$ with R a $d \times k$-matrix and α a fixed k-vector. Now, $\dot{S}(\theta) = R^T$ holds and the estimator from (4.2) reduces to

$$\tilde{\theta}_n = \hat{\theta}_n - \hat{I}_n^{-1} R \left(R^T \hat{I}_n^{-1} R\right)^{-1} R^T \left(\hat{\theta}_n - \alpha\right). \tag{4.9}$$

In terms of a $k \times (k - d)$-matrix L, whose columns span the orthocomplement of $[\dot{S}^T(\theta)] = [R]$, this estimator may be written as

$$\tilde{\theta}_n = \alpha + L\left(L^T \hat{I}_n L\right)^{-1} L^T \hat{I}_n \left(\hat{\theta}_n - \alpha\right) \tag{4.10}$$

according to the Appendix of Susyanto and Klaassen (2017). Note that this estimator attains the asymptotic information bound

$$L\left(L^T I(\theta, G, \mathcal{P}) L\right)^{-1} L^T. \tag{4.11}$$

Comparing formula (4.18) of Susyanto and Klaassen (2017) to (4.10) above we note that the approaches of the present paper and of Susyanto and Klaassen (2017) yield exactly the same estimator in the linear case, although the approaches differ in the general case.

Remark 4.5. The estimators $\tilde{\theta}_n$ and $\hat{\theta}_n$ are efficient within the models $\mathcal{Q}$ and $\mathcal{P}$, respectively. Since $\mathcal{Q}$ is a submodel of $\mathcal{P}$, it is easier to estimate θ within $\mathcal{Q}$ than within $\mathcal{P}$. This is visible in the respective limit distributions by comparing (2.2) and (2.3) to (3.12). The difference between the two limit covariance matrices is $I^{-1}\dot{S}^T(\dot{S}I^{-1}\dot{S}^T)^{-1}\dot{S}I^{-1}$, which is positive semidefinite because of the nonsingularity of the symmetric information matrix I, the maximum rank of $\dot{S}$, and the fact that the inverse of a symmetric positive definite matrix is also symmetric positive definite.

By Theorem 4.1, (3.9), and (3.10) we have

$$\tilde{\theta}_n - \theta_0 = L\left(L^T I L\right)^{-1} L^T I\left(\hat{\theta}_n - \theta_0\right) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{4.12}$$

This means that $\tilde{\theta}_n - \theta_0$ may be viewed as a projection of $\hat{\theta}_n - \theta_0$ into $[L]$, approximately. In other words, $\tilde{\theta}_n$ tends to be closer to the true value θ_0 than $\hat{\theta}_n$ in the metric induced by I.

5. Examples

Our construction of (semi)parametrically efficient estimators will be illustrated in this section by some examples. The first example concerns a constraint on the parameter space of a parametric model that results in a subset of the parameter space that cannot be reparametrized as a regular parametric model. This implies that Susyanto and Klaassen (2017) cannot be applied straight out.

Example 5.1 (*Mean Vector in Ellipsoid*)**.** Let $X_1, \ldots, X_n$ be i.i.d. random vectors with a normal distribution with mean $\theta \in \mathbb{R}^k$ and nonsingular $k \times k$ covariance matrix Σ. Of course, the sample mean $\hat{\theta}_n = \bar{X}_n$ is an efficient estimator of θ within this unconstrained model, and $I = \Sigma^{-1}$ is the Fisher information for θ. So, with $\hat{\Sigma}_n$ the sample covariance matrix $\hat{I}_n = \hat{\Sigma}_n^{-1}$ is a consistent estimator of the Fisher information. Let M be a symmetric positive definite $k \times k$ matrix and let $S(\theta) = \theta^T M\theta - c$ with $c > 0$ be the constraining function, which restricts θ to an ellipsoid. Notice $\dot{S}(\theta) = 2\theta^T M$. As a norm on $\mathbb{R}^k$ we choose $\|\theta\|_M = \sqrt{\theta^T M\theta}$, and we determine an efficient estimator of θ within the constrained model applying Theorem 4.1 with this norm. Using the method of Lagrange multipliers in order to compute $\tilde{\theta}_n$ from (4.2) we see that we have to consider

$$\|\theta_n^* - \zeta\|_M^2 + \lambda\left(\|\zeta\|_M^2 - c\right), \tag{5.1}$$

which leads to

$$\tilde{\theta}_n = \frac{\sqrt{c}}{\|\theta_n^*\|_M}\theta_n^*. \tag{5.2}$$

According to (4.1) we have

$$\theta_n^* = \hat{\theta}_n - \frac{\hat{\theta}_n^T M\hat{\theta}_n - c}{2\hat{\theta}_n^T M\hat{\Sigma}_n M\hat{\theta}_n}\hat{\Sigma}_n M\hat{\theta}_n. \tag{5.3}$$

In case Σ is known, we choose $\hat{\Sigma}_n = \Sigma$. Now, if M is such that the ellipsoid and the inner product induced by M are compatible with the covariance matrix Σ, i.e., $M = \Sigma^{-1}$, then (5.3) simplifies to

$$\theta_n^* = \frac{1}{2}\left(1 + \frac{c}{\hat{\theta}_n^T\Sigma^{-1}\hat{\theta}_n}\right)\hat{\theta}_n, \tag{5.4}$$

which is a (data dependent) multiple of $\hat{\theta}_n$. Consequently, (5.2) becomes then

$$\tilde{\theta}_n = \frac{\sqrt{c}}{\|\hat{\theta}_n\|_{\Sigma^{-1}}}\hat{\theta}_n, \tag{5.5}$$

an intuitively reasonable result.

Example 5.2 (*Part of the Parameter Given*)**.** Let the parameter θ consist of two parts, $\theta_1 \in \mathbb{R}^{k-d}$ and $\theta_2 \in \mathbb{R}^d$, and let $\hat{\theta}_n = (\hat{\theta}_{n1}^T, \hat{\theta}_{n2}^T)^T$ be an efficient estimator of $\theta = (\theta_1^T, \theta_2^T)^T$. We consider the submodel in which θ_2 is known to be equal to $\theta_{0,2}$. This submodel may be defined by the equation $S(\theta) = \theta_2 - \theta_{0,2} = 0$. Now the Jacobian matrix $\dot{S}(\theta)$ equals $(0\, I)$ with

0 the $d \times (k-d)$-matrix of zeros and I the $d \times d$ identity matrix. As in Theorem 4.1 we denote by $\hat{I}_n$ a consistent estimator of the Fisher information matrix $I(\theta, G, \mathcal{P})$. We write both these $k \times k$-matrices and their inverses in block matrix form as follows

$$I(\theta, G, \mathcal{P}) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} \text{ and } \hat{I}_n = \begin{pmatrix} \hat{I}_{n,11} & \hat{I}_{n,12} \\ \hat{I}_{n,21} & \hat{I}_{n,22} \end{pmatrix} \tag{5.6}$$

with I_{22} a $d \times d$-matrix and the other submatrices with corresponding dimensions. The inverses of these matrices are denoted by

$$I^{-1}(\theta, G, \mathcal{P}) = \begin{pmatrix} I_{11\cdot 2}^{-1} & -I_{11\cdot 2}^{-1} I_{12} I_{22}^{-1} \\ -I_{22\cdot 1}^{-1} I_{21} I_{11}^{-1} & I_{22\cdot 1}^{-1} \end{pmatrix} \tag{5.7}$$

and

$$\hat{I}_n^{-1} = \begin{pmatrix} \hat{I}_{n,11\cdot 2}^{-1} & -\hat{I}_{n,11\cdot 2}^{-1} \hat{I}_{n,12} \hat{I}_{n,22}^{-1} \\ -\hat{I}_{n,22\cdot 1}^{-1} \hat{I}_{n,21} \hat{I}_{n,11}^{-1} & \hat{I}_{n,22\cdot 1}^{-1} \end{pmatrix}, \tag{5.8}$$

respectively, with $I_{11\cdot 2} = I_{11} - I_{12} I_{22}^{-1} I_{21}$ and the other $\cdot$-submatrices defined analogously; cf. Section 2.4 of Bickel et al. (1993). According to Theorem 4.1 an efficient estimator of θ for the restricted model with $\theta_2 = \theta_{0,2}$ is

$$\begin{aligned} \theta_n^* &= \hat{\theta}_n - \hat{I}_n^{-1} \begin{pmatrix} 0 \\ I \end{pmatrix} \left((0\, I) \hat{I}_n^{-1} \begin{pmatrix} 0 \\ I \end{pmatrix} \right)^{-1} \left(\hat{\theta}_{n,2} - \theta_{0,2} \right) \\ &= \hat{\theta}_n - \begin{pmatrix} -\hat{I}_{n,11\cdot 2}^{-1} \hat{I}_{n,12} \hat{I}_{n,22}^{-1} \\ \hat{I}_{n,22\cdot 1}^{-1} \end{pmatrix} \hat{I}_{n,22\cdot 1} \left(\hat{\theta}_{n,2} - \theta_{0,2} \right) \\ &= \begin{pmatrix} \hat{\theta}_{n,1} + \hat{I}_{n,11\cdot 2}^{-1} \hat{I}_{n,12} \hat{I}_{n,22}^{-1} \hat{I}_{n,22\cdot 1} \left(\hat{\theta}_{n,2} - \theta_{0,2} \right) \\ \theta_{0,2} \end{pmatrix}, \end{aligned} \tag{5.9}$$

which may be simplified to

$$\begin{aligned} \theta_n^* &= \begin{pmatrix} \hat{\theta}_{n,1} + \hat{I}_{n,11\cdot 2}^{-1} \left[\hat{I}_{n,11} \hat{I}_{n,11}^{-1} \hat{I}_{n,12} - \hat{I}_{n,12} \hat{I}_{n,22}^{-1} \hat{I}_{n,21} \hat{I}_{n,11}^{-1} \hat{I}_{n,12} \right] \left(\hat{\theta}_{n,2} - \theta_{0,2} \right) \\ \theta_{0,2} \end{pmatrix} \\ &= \begin{pmatrix} \hat{\theta}_{n,1} + \hat{I}_{n,11}^{-1} \hat{I}_{n,12} \left(\hat{\theta}_{n,2} - \theta_{0,2} \right) \\ \theta_{0,2} \end{pmatrix}. \end{aligned}$$

Note that θ_n^* satisfies the condition $S(\theta_n^*) = 0$ and consequently (4.2) shows $\tilde{\theta}_n = \theta_n^*$. With $\theta_0 = (\theta_{0,1}^T, \theta_{0,2}^T)^T$ the true parameter value, we obtain by (5.9)

$$\theta_n^* - \theta_0 = \begin{pmatrix} \left(\hat{\theta}_{n,1} - \theta_{0,1} \right) + \hat{I}_{n,11\cdot 2}^{-1} \hat{I}_{n,12} \hat{I}_{n,22}^{-1} \hat{I}_{n,22\cdot 1} \left(\hat{\theta}_{n,2} - \theta_{0,2} \right) \\ 0 \end{pmatrix}. \tag{5.10}$$

Partitioning the score function and the efficient score function for θ within $\mathcal{P}$ (cf. (3.4)) as follows (cf. Section 2.4 of Bickel et al., 1993)

$$\dot{\ell}(\theta_0) = \dot{\ell} = \begin{pmatrix} \dot{\ell}_1 \\ \dot{\ell}_2 \end{pmatrix} \text{ and } \ell^*(\theta_0) = \ell^* = \begin{pmatrix} \ell_1^* \\ \ell_2^* \end{pmatrix} = \begin{pmatrix} \dot{\ell}_1 - I_{12} I_{22}^{-1} \dot{\ell}_2 \\ \dot{\ell}_2 - I_{21} I_{11}^{-1} \dot{\ell}_1 \end{pmatrix}, \tag{5.11}$$

we see that the efficiency of $\hat{\theta}_n$, the consistency of $\hat{I}_n$, and (5.10) show that θ_n^* is asymptotically linear with influence function

$$\begin{aligned} &\begin{pmatrix} I_{11\cdot 2}^{-1} \ell_1^* + I_{11\cdot 2}^{-1} I_{12} I_{22}^{-1} I_{22\cdot 1} I_{22\cdot 1}^{-1} \ell_2^* \\ 0 \end{pmatrix} \\ &\quad = \begin{pmatrix} I_{11\cdot 2}^{-1} \left(\dot{\ell}_1 - I_{12} I_{22}^{-1} \dot{\ell}_2 \right) + I_{11\cdot 2}^{-1} I_{12} I_{22}^{-1} \left(\dot{\ell}_2 - I_{21} I_{11}^{-1} \dot{\ell}_1 \right) \\ 0 \end{pmatrix} \\ &\quad = \begin{pmatrix} \left[I_{11\cdot 2}^{-1} I_{11} I_{11}^{-1} - I_{11\cdot 2}^{-1} I_{12} I_{22}^{-1} I_{21} I_{11}^{-1} \right] \dot{\ell}_1 \\ 0 \end{pmatrix} = \begin{pmatrix} I_{11}^{-1} \dot{\ell}_1 \\ 0 \end{pmatrix}. \end{aligned} \tag{5.12}$$

The last expression of (5.12) is no surprise. Indeed, the influence function $I_{11}^{-1} \dot{\ell}_1$ is efficient for estimating θ_1 within the submodel $\mathcal{Q}$ of $\mathcal{P}$, in which $\theta_{0,2}$ is given. As $\theta_2 = \theta_{0,2}$ is known, the efficient influence function for estimating θ_2 vanishes. Similar computations as above show that (4.3) from Theorem 4.1 also yields the right hand side of (5.12) as efficient influence function.

Next, we will briefly discuss the examples from Section 5 of Susyanto and Klaassen (2017), in light of the results of the present paper. As the constraints in their Example 5.1 (multivariate normal distributions under the constraint of equal means) and their Example 5.3.1 (Gaussian copula model under the constraint of equal correlations) can be formulated via linear equations, Remark 4.4 is applicable and our Theorem 4.1 yields the same results as in their examples. Therefore, these examples will be skipped.

Example 5.3 (*Coefficient of Variation Known*). As in Example 5.2 of Susyanto and Klaassen (2017), we consider the location-scale family corresponding to $g(\cdot)$, an absolutely continuous density on $(\mathbb{R}, \mathcal{B})$ with mean 0, variance 1, and derivative $g'(\cdot)$. Let $X_1, \dots, X_n$, be i.i.d. with density $\sigma^{-1}g((\cdot - \mu)/\sigma)$ and let $\bar{\mu}_n$ and $\bar{\sigma}_n$ be efficient estimators of μ and σ, respectively. We assume that the 2×2 matrix I has well-defined and finite components $I_{11} = \int (g'/g)^2 g$, $I_{12} = I_{21} = \int x(g'/g(x))^2 g(x)dx$, and $I_{22} = \int [xg'/g(x) + 1]^2 g(x)dx$. Then, the Fisher information matrix $I(\theta, G, \mathcal{P})$ as defined in (2.3) with $\theta = (\mu, \sigma)^T$ equals $\sigma^{-2}I$.

The constraint that the coefficient of variation σ/μ equals the given constant c, may be put in the linear form $S(\theta) = c\theta_1 - \theta_2$. By Remark 4.4 and Example 5.1 of Susyanto and Klaassen (2017) this implies that the efficient estimator $\tilde{\theta}_n$ of θ within the constraint model $\mathcal{Q}$ from Theorem 4.1 equals

$$\tilde{\theta}_n = (\hat{\mu}_n, c\hat{\mu}_n)^T \tag{5.13}$$

with

$$\hat{\mu}_n = \left(I_{11} + 2cI_{12} + c^2 I_{22}\right)^{-1} \left[(I_{11} + cI_{12})\,\bar{\mu}_n + (I_{12} + cI_{22})\,\bar{\sigma}_n\right]. \tag{5.14}$$

Similar relations hold for the symmetric and normal cases as discussed in Example 5.1 of Susyanto and Klaassen (2017). Note that one gets another, but still efficient estimator of θ, if one formulates the constraint in a nonlinear way. Choosing e.g. $S(\theta) = \theta_2/\theta_1 - c$, we arrive by Theorem 4.1 at $\theta_n^* = (\mu_n^*, \sigma_n^*)^T$, where straightforward computations with $\bar{c}_n = \bar{\sigma}_n/\bar{\mu}_n$ yield

$$\mu_n^* = \left(I_{11} + 2\bar{c}_n I_{12} + \bar{c}_n^2 I_{22}\right)^{-1} \left[(I_{11} + \{2\bar{c}_n - c\}I_{12})\,\bar{\mu}_n + (I_{12} + \{2\bar{c}_n - c\}I_{22})\,\bar{\sigma}_n\right] \tag{5.15}$$

and

$$\sigma_n^* = \left(I_{11} + 2\bar{c}_n I_{12} + \bar{c}_n^2 I_{22}\right)^{-1} \left[(cI_{11} + c\bar{c}_n I_{12})\,\bar{\mu}_n + \left(\bar{c}_n I_{12} + \bar{c}_n^2 I_{22}\right)\bar{\sigma}_n\right]. \tag{5.16}$$

Indeed, this estimator is asymptotically equivalent to the one from (5.13), but the corresponding coefficient of variation does not equal c. The projection from (4.2) of $\theta_n^* = (\mu_n^*, \sigma_n^*)^T$ yields $\tilde{\theta}_n = (\tilde{\mu}_n, c\tilde{\mu}_n)^T$ with

$$\tilde{\mu}_n = \left(I_{11} + 2\bar{c}_n I_{12} + \bar{c}_n^2 I_{22}\right)^{-1} \left[\left(I_{11} + \frac{2\bar{c}_n - c + c^2\bar{c}_n}{1+c^2} I_{12}\right)\bar{\mu}_n + \left(\frac{1 + c\bar{c}_n}{1+c^2} I_{12} + \frac{2\bar{c}_n - c + c\bar{c}_n^2}{1+c^2} I_{22}\right)\bar{\sigma}_n\right], \tag{5.17}$$

which is asymptotically equivalent to $\hat{\mu}_n$, but also differs from it.

Fig. 1 presents the results of a small simulation study for normal random variables and sample sizes $n = 100, \dots, 1000$. It shows that the performance of $\hat{\mu}_n$ and $\tilde{\mu}_n$ is virtually the same, but improves substantially on the performance of the unconstrained estimator $\bar{\mu}_n$ of μ, which does not use the information about μ contained in $\bar{\sigma}_n$.

Example 5.4 (*Gaussian Copula Models*). With Φ the standard normal distribution function, let

$$\mathbf{X}_1 = (X_{1,1}, \dots, X_{1,m})^T, \dots, \mathbf{X}_n = (X_{n,1}, \dots, X_{n,m})^T$$

be i.i.d. random vectors with marginal distribution functions $F_1, \dots, F_m$, such that

$$(\Phi^{-1}(F_1(X_{i,1})), \dots, \Phi^{-1}(F_m(X_{i,m})))^T, \quad i = 1, \dots, n,$$

have a normal distribution with mean vector 0 and positive definite covariance and hence correlation matrix $C(\theta)$. Here $\theta \in \mathbb{R}^{m(m-1)/2}$ is the parameter vector of interest that summarizes all correlation coefficients ρ_{rs}, $1 \le r < s \le m$. As argued in Susyanto and Klaassen (2017), within this semiparametric starting model

$$\mathcal{P} = \{P_{\theta,G} \,:\, \theta = (\rho_{12}, \dots, \rho_{(m-1)m})^T, G = (F_1(\cdot), \dots, F_m(\cdot)) \in \mathcal{G}\} \tag{5.18}$$

the Van der Waerden or normal scores rank correlation coefficients

$$\hat{\rho}_{rs}^{(n)} = \frac{\frac{1}{n}\sum_{j=1}^n \Phi^{-1}\left(\frac{n}{n+1}\mathbb{F}_r^{(n)}(X_{j,r})\right)\Phi^{-1}\left(\frac{n}{n+1}\mathbb{F}_s^{(n)}(X_{j,s})\right)}{\frac{1}{n}\sum_{j=1}^n \left[\Phi^{-1}\left(\frac{j}{n+1}\right)\right]^2}, \quad 1 \le r < s \le m, \tag{5.19}$$

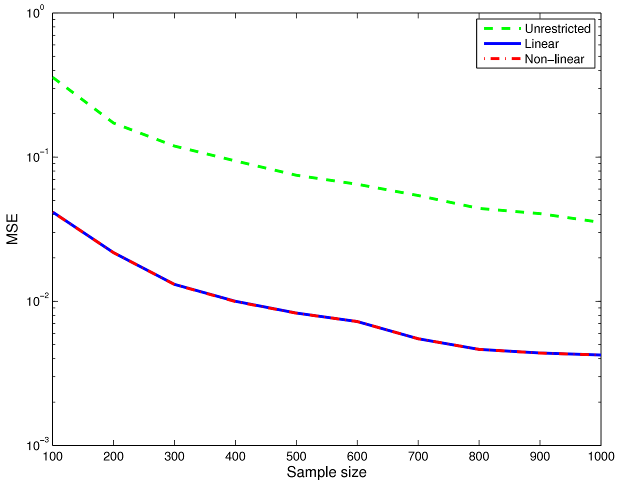

Fig. 1. Normal random variables with $\mu = 3$, $\sigma = 6$, $n = 100, \ldots, 1000$ with step size 100. Each sample has been simulated 1000 times. The mean square error is given for $\bar{\mu}_n$ (green), $\hat{\mu}_n$ (blue), $\tilde{\mu}_n$ (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with $\mathbb{F}_r^{(n)}$ and $\mathbb{F}_s^{(n)}$ being the marginal empirical distributions of F_r and F_s, respectively, constitute a semiparametrically efficient estimator $\hat{\theta}_n$ of θ.

As mentioned before we will not discuss here the Gaussian copula model with equal correlations from Example 5.3.1 of Susyanto and Klaassen (2017), because our Theorem 4.1 yields the same results as in that Example 5.3.1. Therefore, we restrict attention to the other copula example from ibid., namely the four-dimensional circular Gaussian copula model as submodel of $\mathcal{P}$. With $m = 4$ and $\theta = (\rho_{12}, \rho_{14}, \rho_{23}, \rho_{34}, \rho_{13}, \rho_{24})^T$ this submodel can be written as

$$\mathcal{Q} = \left\{ P_{\theta,G} \,:\, \theta = (\rho, \rho, \rho, \rho, \rho^2, \rho^2),\ \rho \in (-1, 1),\ G \in \mathcal{G} \right\} \subset \mathcal{P}. \tag{5.20}$$

(By Sylvester's criterion it may be verified that the four-dimensional circular correlation matrix is positive definite if and only if $|\rho| < 1$ holds. Note that the restriction $-1/3 < \rho < 1$ from Susyanto and Klaassen (2017) is unnecessary.)

To apply Theorem 4.1 we keep Remark 4.1 in mind and choose as in Susyanto and Klaassen (2017)

$$\hat{I}_n^{-1} = I^{-1}(f(\bar{\rho}_n), G, \mathcal{P}) \text{ with } I^{-1}(f(\rho), G, \mathcal{P}) = \frac{1}{2}\left(1 - \rho^2\right)^2 \tag{5.21}$$

$$\begin{pmatrix} 2 & \rho^2 & \rho^2 & 2\rho^2 & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) \\ \rho^2 & 2 & 2\rho^2 & \rho^2 & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) \\ \rho^2 & 2\rho^2 & 2 & \rho^2 & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) \\ 2\rho^2 & \rho^2 & \rho^2 & 2 & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) \\ \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & 2\left(1+\rho^2\right)^2 & 4\rho^2 \\ \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & \rho\left(2+\rho^2\right) & 4\rho^2 & 2\left(1+\rho^2\right)^2 \end{pmatrix}$$

and

$$\begin{aligned} \bar{\rho}_n = \frac{1}{6}\Big\{ &\left[\hat{\theta}_{n1} + \hat{\theta}_{n2} + \hat{\theta}_{n3} + \hat{\theta}_{n4}\right] \\ &+ \left[\sqrt{\hat{\theta}_{n5} \vee 0} + \sqrt{\hat{\theta}_{n6} \vee 0}\right] \operatorname{sign}\left(\hat{\theta}_{n1} + \hat{\theta}_{n2} + \hat{\theta}_{n3} + \hat{\theta}_{n4}\right) \Big\}. \end{aligned} \tag{5.22}$$

We choose the equations $S(\theta) = 0$ that reduce $\mathcal{P}$ to $\mathcal{Q}$, as follows

$$S(\theta) = \begin{pmatrix} \theta_1 + \theta_2 - \theta_3 - \theta_4 \\ \theta_1 - \theta_2 + \theta_3 - \theta_4 \\ \theta_1 - \theta_2 - \theta_3 + \theta_4 \\ 2\theta_5 - \frac{1}{2}[\theta_1^2 + \theta_2^2 + \theta_3^2 + \theta_4^2] \\ \theta_5 - \theta_6 \end{pmatrix} = 0 \tag{5.23}$$

and consequently we have

$$\dot{S}(\theta)=\begin{pmatrix} 1 & 1 & -1 & -1 & 0 & 0 \\ 1 & -1 & 1 & -1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 \\ -\theta_1 & -\theta_2 & -\theta_3 & -\theta_4 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}. \tag{5.24}$$

Again in line with Remark 4.1 we replace in $\dot{S}(\hat{\theta}_n)$ from (4.1) the efficient estimator $\hat{\theta}_n$ within $\mathcal{P}$ as given by (5.19) by $\bar{\theta}_n=(\bar{\rho}_n,\bar{\rho}_n,\bar{\rho}_n,\bar{\rho}_n,\bar{\rho}_n^2,\bar{\rho}_n^2)^T$. This results in

$$\dot{S}(\bar{\theta}_n)=\begin{pmatrix} 1 & 1 & -1 & -1 & 0 & 0 \\ 1 & -1 & 1 & -1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 \\ -\bar{\rho}_n & -\bar{\rho}_n & -\bar{\rho}_n & -\bar{\rho}_n & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \tag{5.25}$$

and

$$\begin{aligned} &\left(\dot{S}(\bar{\theta}_n)\hat{I}_n^{-1}\dot{S}^T(\bar{\theta}_n)\right)^{-1} \\ &= \frac{1}{4}(1-\bar{\rho}_n^2)^{-4}\begin{pmatrix} 1-\bar{\rho}_n^2 & 0 & 0 & 0 & 0 \\ 0 & 1-\bar{\rho}_n^2 & 0 & 0 & 0 \\ 0 & 0 & (1-\bar{\rho}_n^2)^2 & 0 & 0 \\ 0 & 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & -2 & 4-\dfrac{4\bar{\rho}_n^2}{1+\bar{\rho}_n^4} \end{pmatrix}. \end{aligned} \tag{5.26}$$

Subsequently more computation shows that the slightly modified estimator θ_n^* from (4.1) becomes

$$\begin{aligned} \theta_{n1}^* &= \theta_{n2}^*=\theta_{n3}^*=\theta_{n4}^* \\ &= \frac{1}{4}\left[\hat{\rho}_{12}^{(n)}+\hat{\rho}_{14}^{(n)}+\hat{\rho}_{23}^{(n)}+\hat{\rho}_{34}^{(n)}\right] \\ &\quad+\frac{1}{4}\bar{\rho}_n(1-\bar{\rho}_n^2)^{-1}\left[\left(\hat{\rho}_{12}^{(n)2}+\hat{\rho}_{14}^{(n)2}+\hat{\rho}_{23}^{(n)2}+\hat{\rho}_{34}^{(n)2}\right)-2\left(\hat{\rho}_{13}^{(n)}+\hat{\rho}_{24}^{(n)}\right)\right], \\ \theta_{n5}^* &= \theta_{n6}^*=\frac{1}{4}(1+\bar{\rho}_n^2)(1-\bar{\rho}_n^2)^{-1}\left[\hat{\rho}_{12}^{(n)2}+\hat{\rho}_{14}^{(n)2}+\hat{\rho}_{23}^{(n)2}+\hat{\rho}_{34}^{(n)2}\right] \\ &\quad-\bar{\rho}_n^2(1-\bar{\rho}_n^2)^{-1}\left[\hat{\rho}_{13}^{(n)}+\hat{\rho}_{24}^{(n)}\right]. \end{aligned} \tag{5.27}$$

The derivation of the efficient estimator $\tilde{\theta}_n$ of (4.2) from this θ_n^* under the Euclidean norm has to be done numerically as it involves a cubic equation. This has been done in the small simulation study resulting in Fig. 2, which presents the results for sample sizes $n = 100,\ldots,1000$. The mean square error is defined here as the mean of the squared Euclidean distance between the estimator and the estimand. Again, the results show that the performance of our estimator $\tilde{\theta}_n$ and the estimator obtained by Susyanto and Klaassen (2017) is virtually the same, but differs substantially from the performance of the unconstrained estimator $\hat{\theta}_n$ of θ.

Example 5.5 (*Partial Spline Linear Regression*)**.** As in Example 5.4 of Susyanto and Klaassen (2017) the observations are realizations of i.i.d. copies of the random vector $X=(Y,Z^T,U^T)^T$ with the structure

$$Y=\theta^T Z+\psi(U)+\varepsilon, \tag{5.28}$$

where the measurement error ε is independent of Z and U, has mean 0, finite variance, and finite Fisher information for location, and where $\psi(\cdot)$ is a real valued function. Schick (1993) presents an efficient estimator of θ and a consistent estimator of the semiparametric Fisher information matrix in his Theorem 8.1. Consequently our Theorem 4.1 may be applied directly. In the linear case of Remark 4.4 the parameter of interest θ within the submodel $\mathcal{Q}$ may be reparametrized by $\theta=\alpha+L\nu$ with the vector α and the matrix L known. Now ν is the parameter of interest and we return to the situation of (5.28) with $X=(Y-\alpha^T Z, Z^T L, U^T)^T$.

Example 5.6 (*Restricted Maximum Likelihood Estimator*)**.** Most literature on restricted maximum likelihood estimation considers linear restrictions on the parameters and applies iterative procedures; cf. Nyquist (1991), Kim and Taylor (1995) and Jamshidian (2004). The approach as described in Remark 4.4 with $\hat{\theta}_n$ an unrestricted maximum likelihood estimator and the approach as mentioned in Example 5.5 of Susyanto and Klaassen (2017) avoid such iterative procedures. Moreover, our Theorem 4.1 is not constrained to linear restrictions, and neither is Theorem 4.1 of Susyanto and Klaassen (2017).

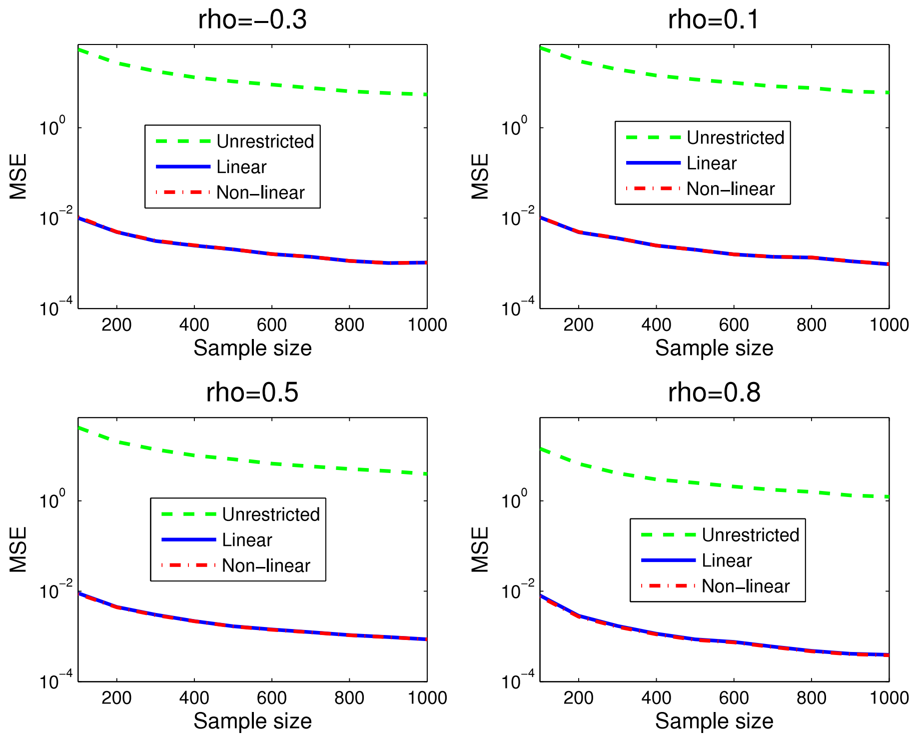

Fig. 2. Four-dimensional circular Gaussian copula random vectors with $\rho = -0.3, 0.1, 0.5, 0.8$, $n = 100, \ldots, 1000$ with step size 100. Each sample has been simulated 1000 times. The mean square error is given for the unrestricted estimator $\hat{\theta}_n$ (green), the estimator obtained from Susyanto and Klaassen (2017) (blue), and the estimator $\tilde{\theta}_n$ (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Conclusion

In this paper, we have shown that the efficient influence function for estimation of θ within the semiparametric model

$$\mathcal{Q} = \left\{P_{\theta,G} \;:\; S(\theta) = 0, \theta \in \Theta, G \in \mathcal{G}\right\}$$

can be obtained by projecting the efficient influence function for estimation of θ within the unconstrained model

$$\mathcal{P} = \left\{P_{\theta,G} \;:\; \theta \in \Theta,\; G \in \mathcal{G}\right\}.$$

It follows that these influence functions are related by

$$\begin{aligned}&\tilde{\ell}(\theta, G, \mathcal{Q})\\ &\quad = \left(J - I^{-1}(\theta, G, \mathcal{P})\dot{S}(\theta)^T \left(\dot{S}(\theta)I^{-1}(\theta, G, \mathcal{P})\dot{S}(\theta)^T\right)^{-1} \dot{S}(\theta)\right) \tilde{\ell}(\theta, G, \mathcal{P})\end{aligned}$$

and hence the corresponding efficient lower bounds by

$$\begin{aligned}&I^{-1}(\theta, G, \mathcal{Q}) = I^{-1}(\theta, G, \mathcal{P})\\ &\quad - I^{-1}(\theta, G, \mathcal{P})\dot{S}(\theta)^T \left(\dot{S}(\theta)I^{-1}(\theta, G, \mathcal{P})\dot{S}(\theta)^T\right)^{-1} \dot{S}(\theta)I^{-1}(\theta, G, \mathcal{P}).\end{aligned}$$

Furthermore, Theorem 4.1 provides a simple method to upgrade an asymptotically efficient estimator for θ within the unconstrained model to an efficient estimator within the constrained model.

Acknowledgments

This research was supported by the Netherlands Organisation for Scientific Research (NWO) via the project Forensic Face Recognition, 727.011.008.

Appendix. Existence of a path with a given direction

Given a continuously differentiable function $S : \Theta \subset \mathbb{R}^k \to \mathbb{R}^d$ with $k > d$. Define

$$\mathcal{M} = \{\theta \in \Theta : S(\theta) = 0\}$$

and let $\theta_0 \in \mathcal{M}$ be such that the Jacobian $\dot{S}(\theta_0)$ of the function $S(\cdot)$ at θ_0 has full-rank d. Suppose that $r \in \mathbb{R}^k$ with $\dot{S}(\theta_0)r = 0$. We would like to construct a path within $\mathcal{M}$ through θ_0 with direction r.

Note that according to the Implicit Function Theorem, there exists an open subset $U \subset \mathbb{R}^{k-d}$, $0 \in U$, and a unique continuously differentiable function $\phi : U \to \mathcal{M}$ with $\phi(0) = \theta_0$ (usually called parametrization). If $\dot{\phi}_0$ denotes the Jacobian of the function ϕ at 0, then the chain rule gives

$$\dot{S}(\theta_0)\dot{\phi}_0 = 0$$

in view of $S(\phi(u)) = 0$ for every $u \in U$. This implies

$$im(\dot{\phi}_0) \subset \dot{S}(\theta_0)^{\perp}.$$

Since $dim(im(\dot{\phi}_0)) = k - d = dim(\dot{S}(\theta_0)^{\perp})$ we obtain

$$im(\dot{\phi}_0) = \dot{S}(\theta_0)^{\perp}.$$

Consequently, the direction r has to belong to $im(\dot{\phi}_0)$, which means that there exists a $\nu \in U$ with $\dot{\phi}_0\nu = r$. Now define a path

$$\{\theta_\eta\} = \{\phi(\eta\nu) \ : \ \eta \in \mathbb{R}, \ |\eta| < \varepsilon\}$$

for sufficiently small $\varepsilon > 0$, which obviously passes through θ_0 because of $\phi(0) = \theta_0$. Then, we have

$$\left|\theta_\eta - \theta_0 - \eta r\right| = |\phi(\eta\nu) - \phi(0) - \eta r| \leq \left|\eta\dot{\phi}_0\nu - \eta r\right| + o(|\eta|) = o(|\eta|).$$

References

Aitchison, J., Silvey, S.D., 1958. Maximum-likelihood estimation of parameters subject to restraints. Ann. Math. Stat. 29 (3), 813–828. http://dx.doi.org/10.1214/aoms/1177706538.

Andrews, D.W.K., 1999. Estimation when a parameter is on a boundary. Econometrica 67 (6), 1341–1383. http://dx.doi.org/10.1111/1468-0262.00082.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y., Wellner, J.A., 1993. Efficient and Adaptive Estimation for Semiparametric Models. In: Johns Hopkins Series in the Mathematical Sciences, The Johns Hopkins University Press. Paperbound (1998), Springer New York.

Broniatowski, M., Keziou, A., 2012. Divergences and duality for estimation and test under moment condition models. J. Statist. Plann. Inference 142 (9), 2554–2573. http://dx.doi.org/10.1016/j.jspi.2012.03.013.

Cam, L., 1986. Asymptotic Methods in Statistical Decision Theory. Springer Series in Statistics, Springer New York, URL: https://books.google.co.id/books?id=gFlHlT8FtOwC.

Gorman, J.D., Hero, A.O., 1990. Lower bounds for parametric estimation with constraints. IEEE Trans. Inform. Theory 36 (6), 1285–1301. http://dx.doi.org/10.1109/18.59929.

Groeneboom, P., Jongbloed, G., Wellner, J.A., 2001. A canonical process for estimation of convex functions: The invelope of integrated Brownian motion $+t^4$. Ann. Statist. 29 (6), 1620–1652. http://dx.doi.org/10.1214/aos/1015345957.

Haberman, S.J., 1984. Adjustment by minimum discriminant information. Ann. Statist. 12 (3), 971–988. http://dx.doi.org/10.1214/aos/1176346715.

Horowitz, J.L., Lee, S., 2017. Nonparametric estimation and inference under shape restrictions. J. Econometrics 201 (1), 108–126. http://dx.doi.org/10.1016/j.jeconom.2017.06.019, URL: http://www.sciencedirect.com/science/article/pii/S0304407617301057.

Jamshidian, M., 2004. On algorithms for restricted maximum likelihood estimation. Comput. Statist. Data Anal. 45 (2), 137–157. http://dx.doi.org/10.1016/S0167-9473(02)00345-6.

Kim, D.K., Taylor, J.M.G., 1995. The restricted EM algorithm for maximum likelihood estimation under linear restrictions on the parameters. J. Amer. Statist. Assoc. 90 (430), 708–716.

Koshevnik, Y.A., Levit, B.Y., 1977. On a non-parametric analogue of the information matrix. Theory Probab. Appl. 21 (4), 738–753. http://dx.doi.org/10.1137/1121087.

Le Cam, L., 1972. Limits of experiments. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics. University of California Press, Berkeley, Calif., pp. 245–261, URL: https://projecteuclid.org/euclid.bsmsp/1200514095.

Levit, B.Y., 1976. On the efficiency of a class of non-parametric estimates. Theory Probab. Appl. 20 (4), 723–740. http://dx.doi.org/10.1137/1120081.

Marchand, E., Strawderman, W.E., 2004. Estimation in restricted parameter spaces: A review. IMS Lecture Notes Monogr. Ser. 45, 21–44, URLhttp://www.jstor.org/stable/4356296.

Marzetta, T.L., 1993. A simple derivation of the constrained multiple parameter Cramer-Rao bound. IEEE Trans. Signal Process. 41 (6), 2247–2249. http://dx.doi.org/10.1109/78.218151.

Müller, U.U., Wefelmeyer, W., 2002. Estimators for models with constraints involving unknown parameters. Math. Methods Statist. 11 (2), 221–235.

Nyquist, H., 1991. Restricted estimation of generalized linear models. J. R. Stat. Soc. Ser. C. Appl. Stat. 40 (1), 133–141.

Schick, A., 1993. On efficient estimation in regression models. Ann. Statist. 21 (3), 1486–1521. http://dx.doi.org/10.1214/aos/1176349269.

Stoica, P., Ng, B.C., 1998. On the Cramér-Rao bound under parametric constraints. IEEE Signal Process. Lett. 5 (7), 177–179. http://dx.doi.org/10.1109/97.700921.

Susyanto, N., Klaassen, C.A.J., 2017. Semiparametrically efficient estimation of constrained Euclidean parameters. Electron. J. Stat. 11 (2), 3120–3140. http://dx.doi.org/10.1214/17-EJS1308.

Van der Vaart, A.W., 1998. Asymptotic statistics. In: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511802256.

Van Eeden, C., 2006. Restricted Parameter Space Estimation Problems: Admissibility and Minimaxity Properties. In: Lecture Notes in Statistics 188, Springer New York.